AFRL-IF-RS-TR-2005-205
Final Technical Report
May 2005

INVESTIGATION OF A SYBR-GREEN-BASED METHOD TO VALIDATE DNA SEQUENCES FOR DNA COMPUTING

SUNY Geneseo

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2005-205 has been reviewed and is approved for publication

APPROVED: /s/
THOMAS E. RENZ
Project Engineer

FOR THE DIRECTOR: /s/
JAMES A. COLLINS, Acting Chief
Advanced Computing Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 2005	**Final** **Jun 04 – Dec 04**

4. TITLE AND SUBTITLE

INVESTIGATION OF A SYBR-GREEN-BASED METHOD TO VALIDATE DNA SEQUENCES FOR DNA COMPUTING

5. FUNDING NUMBERS

G - FA8750-04-2-0218
PE - 62702F
PR - DNAL
TA - IB
WU - TC

6. AUTHOR(S)

Wendy Pogozelski, Salvatore Priore, Matthew Bernard, Anthony Macula

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Research Foundation of the State University of New York
SUNY Geneseo
1 College Circle
Geneseo NY 14454-1401

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/IFTC
525 Brooks Road
Rome NY 13441-4505

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

AFRL-IF-RS-TR-2005-205

11. SUPPLEMENTARY NOTES

AFRL Project Engineer: Thomas E. Renz/IFTC/(315) 330-3423 Thomas.Renz@rl.af.mil

12a. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

This project validated the generation of DNA sequences called a DNA (n,d) code. Fifteen strands of 16 nucleotides each were designed such that a code strand would hybridize only with its reverse-complement and would not cross-hybridize or miss-pair with any other strand in the set. The code was designed using principles from nearest-neighbor studies and the thermodynamics of base stacking. All possible combinations of strands were tested for their potential to miss-pair. Hybridizations of the sequences representing junctions of strands were also tested. Strands were tested for their potential to cross-hybridize by measuring fluorescence over varying temperatures in the presence of SYBR Green I. Strands were tested in pools to show that hybridization between Watson-Crick complements is still thermodynamically favorable even when all strands are present. Nearly all twenty strands and 56 junction sequences showed suitable preference for their direct complement and did not appreciably miss-pair with any other strand in the code. The general properties and modes of binding of SYBR Green I were explored. The experiments indicate that SYBR Green I is positively-charged and binds to DNA predominantly via electrostatic interactions and groove binding. Its fluorescence increases with the number of base pairs in a predictable manner.

14. SUBJECT TERMS

DNA Computing, Cross Hybridization, Fluorescent DNA Detection, Biomolecular Computing

15. NUMBER OF PAGES 21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

ABSTRACT

We are developing a computing basis that employs new algorithmic paradigms for computing so that many discrete mathematical problems can be solved in linear real time. An important first step is the creation and validation of the DNA sequences to be used in coding. It is crucial that DNA strands behave as predicted from their sequence; otherwise computing errors can result. Here we describe the generation of single-stranded DNA sequences called a DNA (n,d) code. Fifteen strands of 16 nucleotides each were designed such that a code strand would hybridize only with its reverse-complement and would not cross-hybridize or mispair with any other strand in the set to any appreciable extent. The code was designed using principles from nearest-neighbor studies and the thermodynamics of base stacking. We tested every possible combination of strands for their potential to mispair. Since the strands are designed for the ultimate purpose of being linked together in a computing problem, we also tested hybridization of the sequences representing junctions of strands. Strands were tested for their potential to cross-hybridize by measuring fluorescence over varying temperatures in the presence of SYBR Green I, a dye whose fluorescence increases exponentially when bound to double-stranded DNA. Strands were tested in pools of all twenty strands to show that hybridization between Watson-Crick complements is still thermodynamically favorable even when all strands are present. Nearly all twenty strands and 56 junction sequences showed suitable preference for their direct complement and did not appreciably mispair with any other strand in the code. Only two strands were eliminated from the set.

To establish confidence with this assay, we also explored the general properties and modes of binding of SYBR Green. The experiments described here, together with results from others, indicate that SYBR Green I is positively-charged and binds to DNA predominantly via electrostatic interactions and groove binding as opposed to intercalation. Its fluorescence increases with an increasing number of base pairs in a predictable manner.

TABLE of CONTENTS

LIST of FIGURES and TABLES

SUMMARY

A DNA code, in biomolecular computing terms, is defined as a collection of single-stranded DNA molecules or oligonucleotides. Several biomolecular computing architectures rely on the binding or hybridization of these strands with their Watson-Crick base-paired complements. For such architectures to be successful, Watson-Crick duplexes must be significantly more stable than all other possible duplexes in the set. Here we describe the generation and validation of a code of 16-mer oligonucleotides, each synthesized according to computer-generated blueprints.

Our validation approach relied on fluorescence measurements in the presence of the organic cyanine dye SYBR Green I. SYBR Green shows exponentially greater fluorescence in the presence of double-stranded DNA as opposed to single-stranded DNA. We previously demonstrated that this technique can be exploited to distinguish between stably-hybridized Watson-Crick duplexes and unstable mismatches (Pogozelski et al, 2004; USAF Technical Report, AFRL-IF-RS-TR-2004-73, http://stinet.dtic.mil)**.**

To determine the affinity of each DNA strand for every other DNA strand in the code, all oligonucleotides in the code were added to a tube and mixed with SYBR Green I. To facilitate comparison, some tubes contained the complement of the strand being tested while other tubes did not. These solutions of DNA strands and SYBR Green were heated to denature DNA strands and then were slowly cooled to allow stable annealing. The strands were then re-heated and cooled while monitoring fluorescence. Fluorescence was measured using an Applied Biosystems 7700 real-time PCR system equipped with a light source, filters, heating and cooling source, fluorescence detector and software. The advantage of this instrument is its 96-well format and programmable software, allowing for rapid screening. Fluorescence for double-stranded DNA was high at the lower temperatures but would decrease as the temperature was raised and denaturation occurred. Watson-Crick-paired duplexes showed high fluorescence; cross-hybridized strands showed low fluorescence, and unpaired strands showed virtually no fluorescence. The software created negative derivative plots of this fluorescence. The peak of the resultant curves represented the melting temperature (Tm), the temperature at which DNA is 50% single-stranded and 50% double-stranded. Tm is a useful parameter in thermodynamic calculations and provides an indication of the stability of a helix.

These experiments were very useful in their ability to identify DNA strands in the original set whose potential to cross-hybridize was too great to be useful in a DNA code. The strands of the code tested in this experiment were found to be largely highly discriminatory in their ability to bind preferentially to their perfect complements and not to other strands in the set.

Despite success with the technique, our ability to interpret results had been limited initially because many aspects of SYBR Green I binding remained elusive. The concentration, structure, and mode of binding were proprietary. Therefore, we set out to find this information experimentally. We observed a mode of binding that was indicative of groove binding rather than intercalation between base pairs and a predictable increase

in fluorescence based on the length of DNA. These findings increased our confidence with the assay.

INTRODUCTION

Previously (Pogozelski et al., 2004) we described the difference between the canonical Watson-Crick base pairs of DNA and the usually less stable mismatches that can also arise when two single strands of DNA come together to form hydrogen bonds and base stacking interactions. The formation of a duplex from single strands is called hybridization and it forms the basis of many biomolecular computing architectures.

In hybridization-based architectures for DNA computing, cross-hybridized duplexes represent errors. It is therefore crucial that DNA sequences be designed so that the formation of a Watson-Crick duplex is far more energetically favorable than the formation of a cross-hybridized duplex. We designed thirty sequences using algorithms (Macula, 2003) in which base stacking interactions were maximized and in which potential pairing was minimized.

DNA helix stability is dependent on several factors. The greatest contribution, according to both mathematical models and empirical verification, is the vertical stacking (mainly π–π interactions) of adjacent base pairs (Borer et al., 1974). Therefore, the identities of the nearest-neighbor bases are crucially important, as they determine this stacking (Freier et al., 1986). The nearest-neighbor model has been extended for heteroduplex stability to include parameters for the interactions that arise with mismatches (Allawi and SantaLucia, 1997; McDowell and Turner, 1996). These models were used in calculating the potential free energies of strand hybridization and sequence design.

Validation that these strands (called oligonucleotides) avoid mispairing is the second step in making certain that DNA molecules will yield accurate computing results. We used a fast and highly automated method employing the dye SYBR Green I and a Sequence Detection System, also known as a Real-time PCR thermalcycler. This instrument contains a light source, various filters, a 96-well platform, a programmable heating and cooling apparatus, and a fluorescence detector capable of monitoring seven absorption and emission wavelengths. SYBR Green I is a asymmetric positively-charged cyanine dye whose fluorescence emission at 510-520 nm increases markedly in the presence of double-stranded DNA. Recently, a structure was proposed for SYBR Green on the basis of nuclear magnetic resonance (NMR) and mass spectrometry (MS) experiments. The chemical name, which had been proprietary, was reported to be [2-[N-(3-dimethylaminopropyl)-N-propylamino]-4-[2,3-dihydro-3-methyl-(benzo-1,3-thiazol-2-yl)-methylidene]-1-phenyl-quinolinium. (See Figure 1).

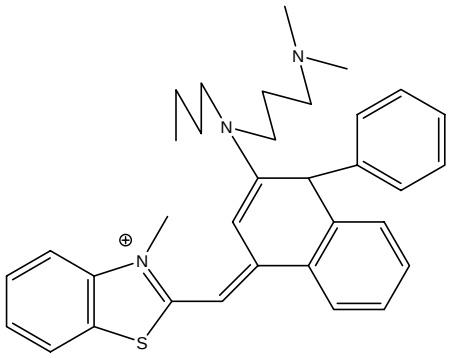

Figure 1. Structure of SYBR Green I [2-[N-(3-dimethylaminopropyl)-N-propylamino]-4-[2,3-dihydro-3-methyl-(benzo-1,3-thiazol-2-yl)-methylidene]-1-phenyl-quinolinium proposed on the basis of Gas Chromatograhpy (GC) and NMR experiments (Zipper et al., 2004).

METHODS

Experiments to Probe SYBR Green Binding

An experiment was designed that would show how fluorescence increases with increasing numbers of base pairs and how fluorescence would be affected by temperature. Five duplexes were created from oligonucleotides, all with the same sequence but differing in length: $C_{12}G_{12}$, $C_{16}G_{16}$, $C_{20}G_{20}$, $C_{18}G_{18}$ and $C_{24}G_{24}$. The strands were synthesized using phosphoramidite chemistry (InVitrogen) and were desalted. No additional purification steps were taken, since we observed no difference between desalted strands and HPLC-purified strands. The lyophilized oligonucleotides were dissolved in 10 mM Tris buffer/1mM EDTA, pH 8, for a concentration of 1 μg/μL (0.48 M). All water used in buffers and dilutions was distilled and deionized using a Millipore purification system. Conditions were the same as for the mispairing experiments described above. SYBR Green I (Applied Biosystems) was supplied as a concentrated stock with no molecular weight or molar concentration data provided. Efforts to obtain this information from Applied Biosystems were unsuccessful but it is likely that the concentration was approximately 10 mg/mL (Zipper et al., 2004). In the absence of hard concentration data, the optimum amount of SYBR Green I was determined empirically. Each well consisted of 0.5 μg of each oligonucleotide, 1X SYBR Green I Master Mix (Applied Biosystems), and enough distilled deioinized water for a 50 μL volume. The Master Mix included a passive reference for standardization of the fluorescence. It was important to keep the concentration of SYBR Green constant, as excess SYBR Green can quench the fluorescence signal (Lipsky et al., 2001). Graphs of fluorescence vs. temperature were created.

Another experiment was conducted to determine whether or not SYBR Green bound to DNA via intercalation. Aliquots of approximately 10 mg supercoiled plasmid (pBR322; Sigma) were treated with 0.1X, 1X, 10X, 100X, 1000X and 10000X SYBR Green and were separated on a 1% agarose gel in 1X Tris-Acetate-EDTA buffer along with an untreated control and controls treated with a restriction enzyme.

Creation of DNA Strands

Sequences for a set of 15 primary DNA strands 16 nucleotides long (strands S1-S15) and their 15 Watson-Crick complement (strands C1-C15) were generated using A.Macula's computing methods. This list is shown in Table I. Sequences representing complements to junctions – that is, half of one strand and half of another, were also generated. For example, strand S1/C2 represents a 32-base product that would result if S1 and C2 were ligated together. These 56 oligonucleotides of 32 bases each are shown in Table II.

The DNA oligonucleotides were also synthesized using phosphoramidite chemistry and were desalted (InVitrogen). Lyophilized oligonucleotides were dissolved in 10 mM Tris buffer/1 mM EDTA for a concentration of 1 μg/μL (0.48 M). All water used in dilutions and buffers was distilled and deionized via a Millipore purification system.

Each well consisted of 0.5 μg of each oligonucleotide, 1X SYBR Green I Master Mix (Applied Biosystems), and enough distilled deioinized water for a 50 μL volume. The Master Mix included a passive reference for standardization of the fluorescence.

Fluorescence Measurements

Fluorescence measurements were made on an Applied Biosystems Model 7000 Sequence Detection System. Each test required set-up of several wells of the 96-well plate. One well contained all sequences except the perfect reverse complement. Another well contained all the strands including the reverse complement. In this way, we could compare the fluorescence obtained for a perfect Watson-Crick duplex with that obtained for mismatches. Furthermore, testing the strands without a perfect complement was an extra-stringent test. We would expect more mismatches in the absence of the competing direct complement.

Immediately prior to insertion into the sequence detection system, strands were heated to 90 °C in a standard thermalcycler to denature all secondary structure. Samples were then slowly cooled to 25 °C in the thermalcycler so that duplexes would form. Upon reaching 25 °C, samples were inserted into the sequence detection system.

Fluorescence emission was monitored at 520 nm using the instrument detector and software over a 35 ° temperature window. Measurements were made by slowly increasing the temperature to 60-70 °C over a period of several minutes. The software converted raw fluorescence data (relative to the passive reference) into melting curves by plotting the negative derivative for fluorescence vs. temperature (-dF/dT vs. T). Data were exported to Microsoft Excel for additional analysis. The maximum of each

derivative curve corresponds to the melting temperature (T_m) of the duplex. The T_m is defined as the temperature at which the DNA is 50% double-stranded.

Table I. DNA Code

***S1 [A-G-G-C-T-A-A-A-G-T-T-A-T-C-A-C]**
*C1 [G-T-G-A-T-A-A-C-T-T-T-A-G-C-C-T]
S2 [G-T-C-T-T-C-G-T-T-T-T-T-T-C-A]
C2 [T-G-A-A-A-A-A-A-C-G-A-A-G-A-C]
S3 [G-C-A-A-G-C-G-A-C-C-A-A-T-A-C-T]
C3 [A-G-T-A-T-T-G-G-T-C-G-C-T-T-G-C]
S4 [T-A-C-C-T-T-T-T-C-T-C-G-A-C-G-C]
C4 [G-C-G-T-C-G-A-G-A-A-A-A-G-G-T-A]
S5 [C-T-C-A-A-T-A-A-A-A-T-G-C-G-C-G]
C5 [C-G-C-G-C-A-T-T-T-T-A-T-T-G-A-G]
S6 [C-G-T-T-G-C-A-C-T-C-A-A-G-A-T-C]
C6 [G-A-T-C-T-T-G-A-G-T-G-C-A-A-C-G]
S7 [G-A-C-T-G-G-A-A-T-G-T-T-T-T-G-T]
C7 [A-C-A-A-A-A-C-A-T-T-C-C-A-G-T-C]
S8 [G-G-A-T-G-C-A-G-G-T-T-G-A-T-T-A]
C8 [T-A-A-T-C-A-A-C-C-T-G-C-A-T-C-C]
S9 [A-A-G-C-C-T-T-A-G-A-A-G-A-G-A-G]
C9 [C-T-C-T-C-T-T-C-T-A-A-G-G-C-T-T]
S10 [T-T-T-C-T-G-T-G-G-C-A-C-T-G-G-T]
C10 [A-C-C-A-G-T-G-C-C-A-C-A-G-A-A-A]
S11 [T-G-T-G-T-G-T-C-C-G-A-T-G-A-G-A]
C11 [T-C-T-C-A-T-C-G-G-A-C-A-C-A-C-A]
S12 [T-T-A-A-A-A-G-A-C-G-T-T-G-G-T-T]
C12 [A-A-C-C-A-A-C-G-T-C-T-T-T-T-A-A]
S13 [T-A-C-G-C-T-A-A-T-C-G-G-T-A-A-G]
C13 [C-T-T-A-C-C-G-A-T-T-A-G-C-G-T-A]
S14 [T-G-G-A-G-G-A-A-C-T-A-C-C-G-G-A]
C14 [T-C-C-G-G-T-A-G-T-T-C-C-T-C-C-A]
S15 [C-C-A-T-A-G-C-T-G-A-G-T-T-C-T-T]
C15 [A-A-G-A-A-C-T-C-A-G-C-T-A-T-G-G]
***S indicates primary strand; C indicates complementary strand**

Table II. Junction Strands

StrandComb[1/2] :=
S1/S2
[A-G-G-C-T-A-A-A-G-T-T-A-T-C-A-C-G-T-C-T-T-C-G-T-T-T-T-T-T-T-C-A]
S1/C2
[A-G-G-C-T-A-A-A-G-T-T-A-T-C-A-C-T-G-A-A-A-A-A-A-A-C-G-A-A-G-A-C]
C1/S2
[G-T-G-A-T-A-A-A-C-T-T-T-A-G-C-C-T-G-T-C-T-T-C-G-T-T-T-T-T-T-T-C-A]
C1/C2
[G-T-G-A-T-A-A-A-C-T-T-T-A-G-C-C-T-T-G-A-A-A-A-A-A-A-C-G-A-A-G-A-C]

StrandComb[2/3] :=
S2/S3
[G-T-C-T-T-C-G-T-T-T-T-T-T-T-C-A-G-C-A-A-G-C-G-A-C-C-A-A-T-A-C-T]
S2/C3
[G-T-C-T-T-C-G-T-T-T-T-T-T-T-C-A-A-G-T-A-T-T-G-G-T-C-G-C-T-T-G-C]
C2/S3
[T-G-A-A-A-A-A-A-A-C-G-A-A-G-A-C-G-C-A-A-G-C-G-A-C-C-A-A-T-A-C-T]
C2/C3
[T-G-A-A-A-A-A-A-A-C-G-A-A-G-A-C-A-G-T-A-T-T-G-G-T-C-G-C-T-T-G-C]

StrandComb[3/4] :=
S3/S4
[G-C-A-A-G-C-G-A-C-C-A-A-T-A-C-T-T-A-C-C-T-T-T-T-C-T-C-G-A-C-G-C]
S3/C4
[G-C-A-A-G-C-G-A-C-C-A-A-T-A-C-T-G-C-G-T-C-G-A-G-A-A-A-A-G-G-T-A]
C3/S4
[A-G-T-A-T-T-G-G-T-C-G-C-T-T-G-C-T-A-C-C-T-T-T-T-C-T-C-G-A-C-G-C]
C3/C4
[A-G-T-A-T-T-G-G-T-C-G-C-T-T-G-C-G-C-G-T-C-G-A-G-A-A-A-A-G-G-T-A]

StrandComb[4/5] :=
S4/S5
[T-A-C-C-T-T-T-T-C-T-C-G-A-C-G-C-C-T-C-A-A-T-A-A-A-A-T-G-C-G-C-G]
S4/C5
[T-A-C-C-T-T-T-T-C-T-C-G-A-C-G-C-C-G-C-G-C-A-T-T-T-T-A-T-T-G-A-G]
C4/S5
[G-C-G-T-C-G-A-G-A-A-A-A-G-G-T-A-C-T-C-A-A-T-A-A-A-A-T-G-C-G-C-G]
C4/C5
[G-C-G-T-C-G-A-G-A-A-A-A-G-G-T-A-C-G-C-G-C-A-T-T-T-T-A-T-T-G-A-G]

StrandComb[5/6] :=
S5/S6
[C-T-C-A-A-T-A-A-A-A-T-G-C-G-C-G-C-G-T-T-G-C-A-C-T-C-A-A-G-A-T-C]
S5/C6
[C-T-C-A-A-T-A-A-A-A-T-G-C-G-C-G-G-A-T-C-T-T-G-A-G-T-G-C-A-A-C-G]
C5/S6
[C-G-C-G-C-A-T-T-T-T-A-T-T-G-A-G-C-G-T-T-G-C-A-C-T-C-A-A-G-A-T-C]
C5/C6
[C-G-C-G-C-A-T-T-T-T-A-T-T-G-A-G-G-A-T-C-T-T-G-A-G-T-G-C-A-A-C-G]

StrandComb[6/7] :=
S6/S7
[C-G-T-T-G-C-A-C-T-C-A-A-G-A-T-C-G-A-C-T-G-G-A-A-T-G-T-T-T-T-G-T]
S6/C7
[C-G-T-T-G-C-A-C-T-C-A-A-G-A-T-C-A-C-A-A-A-A-C-A-T-T-C-C-A-G-T-C]
C6/S7
[G-A-T-C-T-T-G-A-G-T-G-C-A-A-C-G-G-A-C-T-G-G-A-A-T-G-T-T-T-T-G-T]
C6/C7
[G-A-T-C-T-T-G-A-G-T-G-C-A-A-C-G-A-C-A-A-A-A-C-A-T-T-C-C-A-G-T-C]

StrandComb[7/8] :=
S7/S8
[G-A-C-T-G-G-A-A-T-G-T-T-T-T-G-T-G-G-A-T-G-C-A-G-G-T-T-G-A-T-T-A]
S7/C8
[G-A-C-T-G-G-A-A-T-G-T-T-T-T-G-T-T-A-A-T-C-A-A-C-C-TG-C-A-T-C-C]
C7/S8
[A-C-A-A-A-A-C-A-T-T-C-C-A-G-T-C-G-G-A-T-G-C-A-G-G-T-T-G-A-T-T-A]
C7/C8
[A-C-A-A-A-A-C-A-T-T-C-C-A-G-T-C-T-A-A-T-C-A-A-C-C-T-G-C-A-T-C-C]

StrandComb[8/9] :=
S8/S9
[G-G-A-T-G-C-A-G-G-T-T-G-A-T-T-A-A-A-G-C-C-T-T-A-G-A-A-G-A-G-A-G]
S8/C9
[G-G-A-T-G-C-A-G-G-T-T-G-A-T-T-A-C-T-C-T-C-T-T-C-T-A-A-G-G-C-T-T]
C8/S9
[T-A-A-T-C-A-A-C-C-T-G-C-A-T-C-C-A-A-G-C-C-T-T-A-G-A-A-G-A-G-A-G]
C8/C9
[T-A-A-T-C-A-A-C-C-T-G-C-A-T-C-C-C-T-C-T-C-T-T-C-T-A-A-G-G-C-T-T]

StrandComb[9/10] :=
S9/S10
[A-A-G-C-C-T-T-A-G-A-A-G-A-G-A-G-T-T-T-C-T-G-T-G-G-C-A-C-T-G-G-T]
S9/C10
[A-A-G-C-C-T-T-A-G-A-A-G-A-G-A-G-A-C-C-A-G-T-G-C-C-A-C-A-G-A-A-A]
 C9/S10
[C-T-C-T-C-T-T-C-T-A-A-G-G-C-T-T-T-T-T-C-T-G-T-G-G-C-A-C-T-G-G-T]
C9/C10
[C-T-C-T-C-T-T-C-T-A-A-G-G-C-T-T-A-C-C-A-G-T-G-C-C-A-C-A-G-A-A-A]

StrandComb[10/11] :=
S10/S11
[T-T-T-C-T-G-T-G-G-C-A-C-T-G-G-T-T-G-T-G-T-G-T-C-C-G-A-T-G-A-G-A]
S10/C11
[T-T-T-C-T-G-T-G-G-C-A-C-T-G-G-T-T-C-T-C-A-T-C-G-G-A-C-A-C-A-C-A]
C10/S11
[A-C-C-A-G-T-G-C-C-A-C-A-G-A-A-A-T-G-T-G-T-G-T-C-C-G-A-T-G-A-G-A]
C10/C11
[A-C-C-A-G-T-G-C-C-A-C-A-G-A-A-A-T-C-T-C-A-T-C-G-G-A-C-A-C-A-C-A]

StrandComb[11/12] :=
S11/S12
[T-G-T-G-T-G-T-C-C-G-A-T-G-A-G-A-T-T-A-A-A-A-G-A-C-G-T-T-G-G-T-T]
S11/C12
[T-G-T-G-T-G-T-C-C-G-A-T-G-A-G-A-A-A-C-C-A-A-C-G-T-C-T-T-T-T-A-A]
C11/S12
[T-C-T-C-A-T-C-G-G-A-C-A-C-A-C-A-T-T-A-A-A-A-G-A-C-G-T-T-G-G-T-T]
C11/C12

[T-C-T-C-A-T-C-G-G-A-C-A-C-A-C-A-A-A-C-C-A-A-C-G-T-C-T-T-T-T-A-A]

StrandComb[12/13] :=
S12/S13
[T-T-A-A-A-A-G-A-C-G-T-T-G-G-T-T-T-A-C-G-C-T-A-A-T-C-G-G-T-A-A-G]
S12/C13
[T-T-A-A-A-A-G-A-C-G-T-T-G-G-T-T-C-T-T-A-C-C-G-A-T-T-A-G-C-G-T-A]
C12/S13
[A-A-C-C-A-A-C-G-T-C-T-T-T-T-T-A-A-T-A-C-G-C-T-A-A-T-C-G-G-T-A-A-G]
C12/C13
[A-A-C-C-A-A-C-G-T-C-T-T-T-T-T-A-A-C-T-T-A-C-C-G-A-T-T-A-G-C-G-T-A]

StrandComb[13/14] :=
S13/S14
[T-A-C-G-C-T-A-A-T-C-G-G-T-A-A-G-T-G-G-A-G-G-A-A-C-T-A-C-C-G-G-A]
S13/C14
[T-A-C-G-C-T-A-A-T-C-G-G-T-A-A-G-T-C-C-G-G-T-A-G-T-T-C-C-T-C-C-A]
C13/S14
[C-T-T-A-C-C-G-A-T-T-A-G-C-G-T-A-T-G-G-A-G-G-A-A-C-T-A-C-C-G-G-A]
C13/C14
[C-T-T-A-C-C-G-A-T-T-A-G-C-G-T-A-T-C-C-G-G-T-A-G-T-T-C-C-T-C-C-A]

Strand Combination[14/15]
S14/S15
[T-G-G-A-G-G-A-A-C-T-A-C-C-G-G-A-C-C-A-T-A-G-C-T-G-A-G-T-T-C-T-T]
S14/C15
[T-G-G-A-G-G-A-A-C-T-A-C-C-G-G-A- A-A-G-A-A-C-T-C-A-G-C-T-A-T-G-G]
C14/S15
[T-C-C-G-G-T-A-G-T-T-C-C-T-C-C-A-C-C-A-T-A-G-C-T-G-A-G-T-T-C-T-T]
C14/C15
[T-C-C-G-G-T-A-G-T-T-C-C-T-C-C-A-A-A-G-A-A-C-T-C-A-G-C-T-A-T-G-G]

RESULTS and DISCUSSION

Experiments to Probe SYBR Green Binding

To study the mode of binding of SYBR Green to DNA, various experiments were conducted. The first looked at whether or not SYBR Green bound to DNA via intercalation. Because the dye contains aromatic groups, we considered that these rings could stack between base pairs the way that ethidium bromide, another fluorescent dye, binds to DNA. This question was investigated by studying the effect of SYBR Green on the mobility of circular supercoiled plasmid DNA in an agarose electrophoresis gel. Supercoiled DNA, since it has a smaller radius, migrates through agarose more quickly than relaxed circular DNA. If intercalation is predominant, the SYBR Green should cause the plasmid DNA to become overwound or more positively-supercoiled. Since plasmid is naturally negatively-supercoiled, the SYBR Green should initially cancel out the negative supercoils, causing the DNA to migrate more slowly, as it becomes increasingly relaxed. Once the negative supercoils are canceled out, additional SYBR Green molecules cause the DNA to become positively-supercoiled and migrate more quickly.

We did not see evidence of intercalation. All the DNA bands migrated with equal mobility regardless of the amount of ethidium bromide present. These bands showed the same mobility as untreated supercoiled plasmid. There was also no change in the amount

of supercoiled, relaxed, and linear DNA. These results corroborate the work of Zipper et al. who, subsequent to our work, showed that binding of SYBR Green to DNA occurs by groove interactions rather than intercalation (Zipper et al., 2004).

We did make an interesting observation: the SYBR Green migrated to the cathode (the negatively-charged anode) of the cell, indicating that the species is positively-charged. Again, these observations were supported by the work of Zipper et al. who published a structure that showed a positively-charged cyanine dye.

We next needed to determine the dependence of our fluorescence measurements on the number of base pairs present. We were uncertain, for example, if fluorescence depended linearly on the length of the DNA, or if any dependence existed at all. Therefore, we treated a series of duplexes of the same sequence (polyC•polyG) but of various lengths with SYBR Green and measured the fluorescence. We obtained plots of fluorescence vs. temperature and then used the data to create the plot show in Figure 4. Figure 4 shows that fluorescence does depend somewhat on the length of the DNA. At about 75 °, this dependence is linear. At other temperatures, the dependence is not linear, but follows a predictable increasing relationship. The graph also shows that fluorescence is dependent on temperature, and that regardless of the length of the sequence, fluorescence is very low as long as the temperature is below the Tm of the duplex.

Use of Fluorescence to Monitor Hybridization

SYBR Green I shows greatly increased fluorescence when bound to double-stranded DNA. A typical plot of the change in fluorescence as a function of temperature is shown in Figure 2. The sequence G_{16} is being tested for hybridization with C_{16} as well as against the other 24 sequences of the set. Each curve represents a different pair combination of sequences. The magnitude of the change in fluorescence for G_{16} binding to C_{16} is far greater than that of any other combination. Moreover- the temperature which corresponds to the maximum in the curve represents the melting temperature of T_m for this duplex. The T_m is defined as the temperature at which the DNA is 50% single-stranded and 50% double-stranded. This parameter correlates with the thermal stability of the duplex.

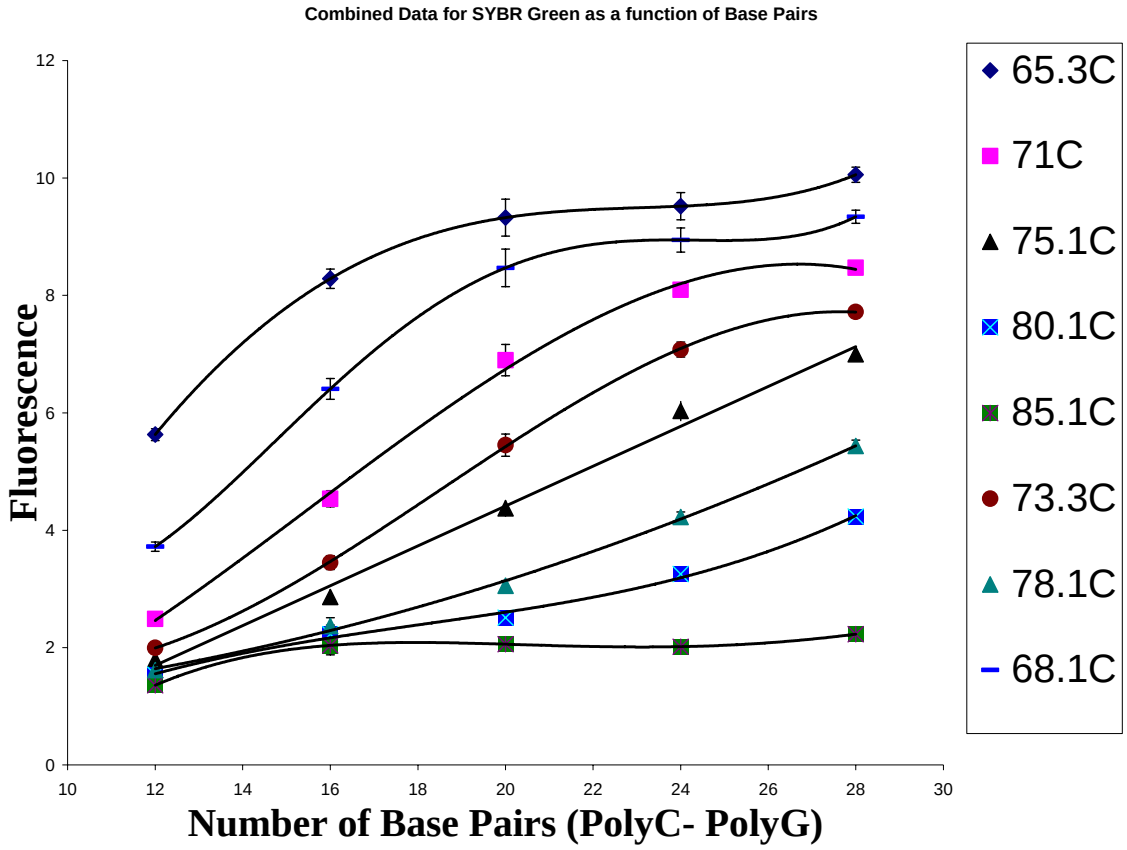

Figure 2. Dependence of fluorescence on duplex length.

We then tested all twenty strands of the DNA code for the ability to distinguish between their perfect complement and any other strands present in the code such that future computing experiments would not be tainted by errors from cross-hybridized strands. None of the strands were found to cross-hybridize.

Since these individual strands will ultimately be linked together to form much longer strands, we also needed to test whether the junctions of these sequences might be able to mispair. For example, while strand S1 might not hybridize to any other strand in the code except its perfect complement C1, we needed to consider whether or not S1S2 (strands S1 and S2 ligated together) would create a new sequence at their junction that would have affinity for other sequences in the code. Therefore, we created 56 strands of 32 nucleotides in length that represented junctions of sequences. For example, Junction S1/C2 represented the ligation product of strand S1 with strand C2. We studied the fluorescence of these junction strands both in the presence and absence of their complements. The complement would be expected to bind to only half of the junction strand. For example, if testing junction strand S1/C2, we would expect strand C1 (the complement of S1) to bind to the first (5') half of the junction, and the strand S2 (the complement of C2) to bind to the second half of the junction. See Figure 3.

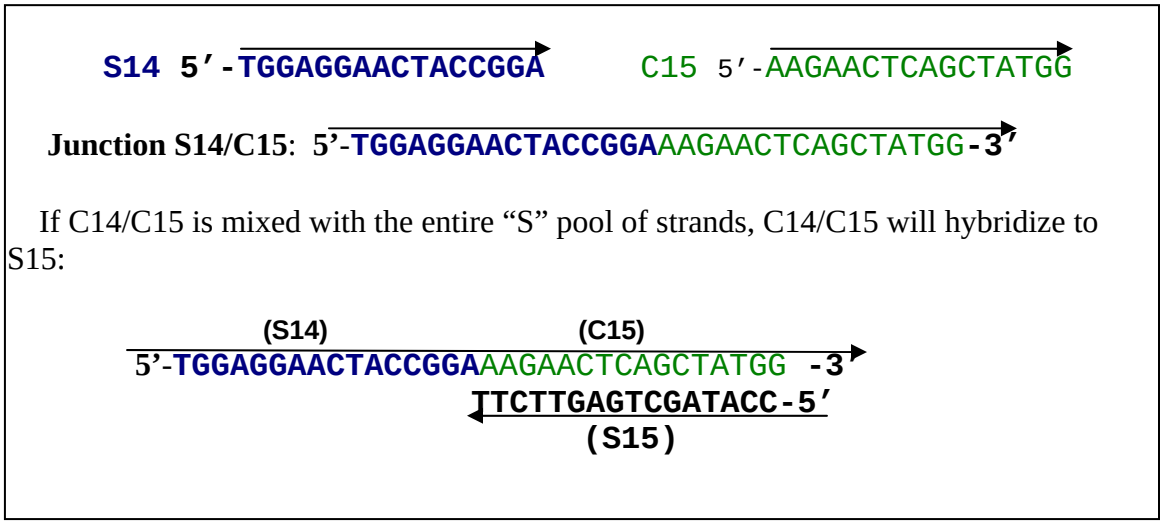

Figure 3. Example of a Junction Strand (S14/C15) composed from strands S14 and C15 and its hybridization to complementary strand S15.

Typical results for a successful experiment to test junction sequences are shown in Figure 4. The graph shows the junction sequence S14/C15 being tested against all other strands in the code. All of the primary strands (S1-S15) and complementary strands (C1-C15) have baseline fluorescence when pooled, indicating that there is no appreciable duplex formation among these strands. When the junction sequence S14/C15 (bearing half of S14 and half of C15 as illustrated in Figure 3) is added to the pool of strands C1-C15, there is no additional fluorescence, indicating that the junction strand does not cross-hybridize to any strand in this set. In mixing the junction sequence S14/C15 with the complement "S" pool of strands (S1-S15), we had to be a little more careful and take extra steps. We knew that junction sequence S14/C15 would hybridize with C14 and S15. Therefore, we tested binding to the "S" pool both in the presence and absence of these complements. Fluorescence was barely above baseline when all the "S" strands (minus S14 and S15) were mixed with the junction C14/C15 and rose dramatically when S15 was added. This is the behavior expected for sequences that do not cross-hybridize. Therefore, these sequences appear to be well-behaved enough to be used in DNA computing.

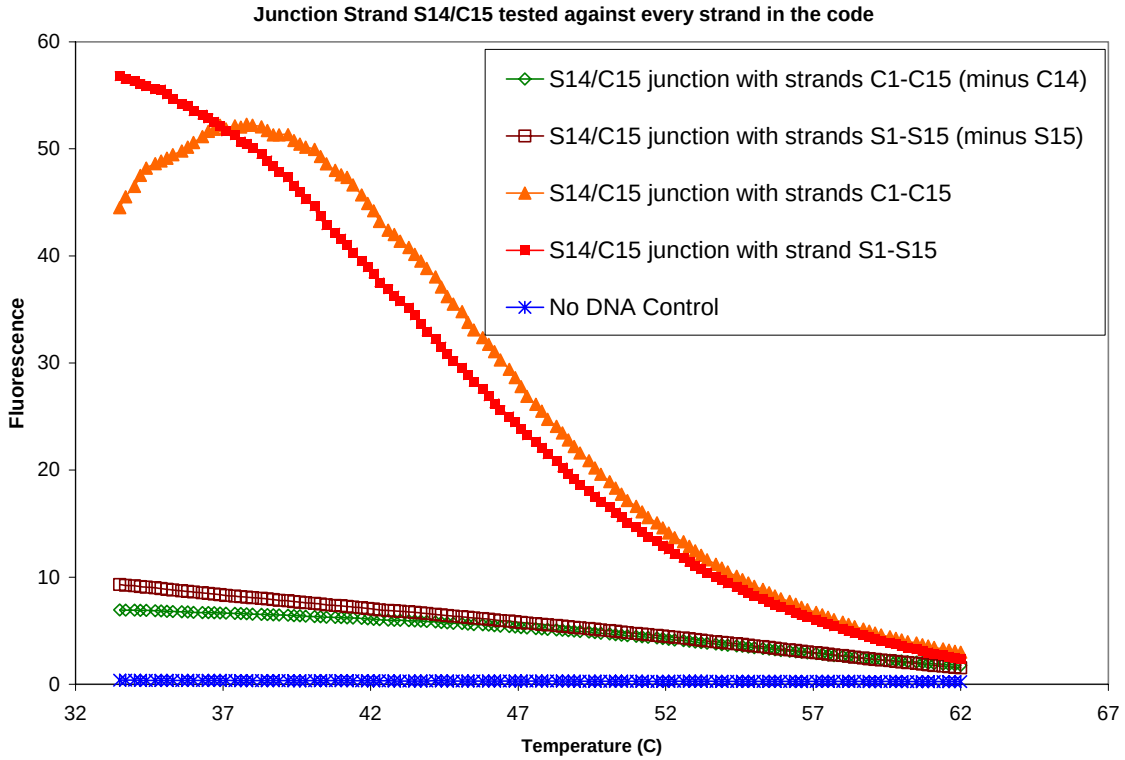

Figure 4. Example of fluorescence experiment testing a junction sequence against all other strands in the DNA code. Closed squares and triangles show fluorescence of the duplex created when half the complementary strand is added. The open symbols and the circle show fluorescence when known complements are omitted.

Most of the fluorescence experiments showed results such as are seen in Figure 4. A few junction strands did exhibit duplex formation with other stand(s) in the code, however. An example is shown in Figure 5.

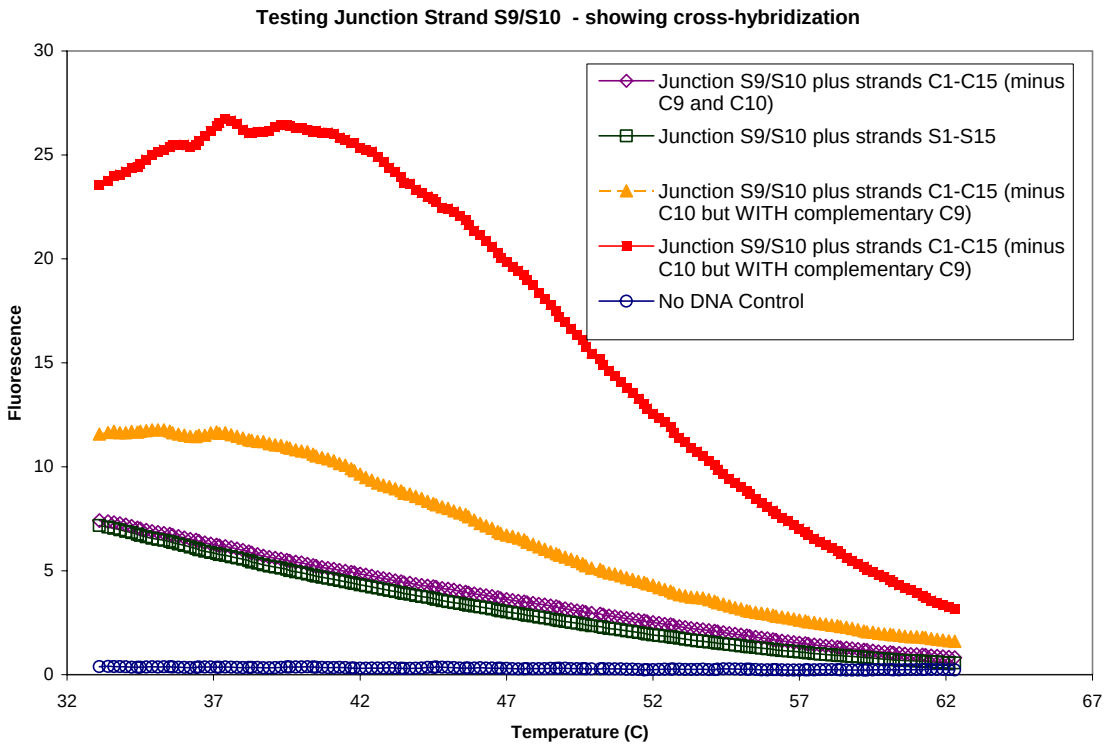

Figure 5. Example of a junction strand being tested against all other strands in the DNA code and showing potential cross-hybridization. The closed squares and triangles (the topmost curves) show fluorescence of the junction sequence when the known complement to half the strand is included. The open circle and square show fluorescence when known complements are not included.

Figure 5, in which the junction S9/S10 was tested against all other strands, shows potential cross-hybridization at low temperatures between the junction sequence and other strands present because fluorescence is more than half that seen when the complement is added. Although the overall fluorescence is low for these strands that are not expected to hybridize, the fluorescence is large when compared to fluorescence when the complement is added. Thus, either S9 or S10 is unsuitable for DNA computing. We eliminated both strands from the code.

It is interesting to consider why one set of duplexes shows far greater fluorescence than the other. The duplex bound to the 5'-side of the top strand (S9/S10•C9) shows much less fluorescence than the duplex bound to the 3'-side (S9/S10•C10). The difference likely reflects a sequence dependence of the SYBR Green.

CONCLUSIONS

The use of SYBR Green along with the Sequence Detection System to predict cross-hybridization of DNA strands is fast and easy and the method is robust. Newly acquired knowledge of the mode of binding and the structure assists in data interpretation. The use of the Sequence Detection System also allows for the processing of large numbers of samples including junction strands. Thirteen out of the fifteen strands in this code were found to be sufficiently selective to be used in hybridization-based computing methods.

REFERENCES

Allawi, H.T., SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G-T mismatches in DNA. Biochemistry, 36, 10581-10594.

Andronescu, M.; Aguirre,Hernández, R.; Condon, A.; Hoose, H.H. (2003) RNAsoft: a suite of RNA secondary structure prediction and design software tools. Nucleic Acids Research, 31, 13, 3416-3422.

Bengtsson, M., Karlsson, H.J., Westman, G. and Kubista, M. (2003) A new minor groove binding asymmetric cyanine reporter dye for real-time PCR. Nucleic Acids Research, 31(8), e45.

Borer, P.N., Gengler, B., Tinoco, I., Jr. (1974) Stability of ribonucleic acid double-stranded helices. J. Mol. Biol. 86, 843-853.

Freier, S.M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R. et al. (1986). Stability of XGCGCp- GCGCYp- and XGCGCYp helixes: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. Biochemistry, 25- 3214-3219.

Lipsky, R.H., Mazzanti, C.M., Rudolph, J.G., Xu, K., Vyas, G., Bozak, D., Radel, M.Q., Goldman, D. (2001) DNA melting analysis for detection of single nucleotide polymorphisms, Clinical Chemistry, 47, 635-644.

Macula, A.J. (2003) US Air Force Office of Scientific Research Technical Report, DNA TAT Codes TR-2003-57- AFRL-IF-RS- http://stinet.dtic.mil/cgi-bin/fulcrum_main.pl

McDowell, J.A., Turner, D.H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of $(rGAGGUCUC)_2$ by two-dimensional NMR and simulated annealing. Biochemistry, 35, 14077-14089.

Pogozelski, W.K., Bernard, M.P. and Macula, A. (2004) DNA code validation using experimental fluorescence measurements and thermodynamic calculations. ARFL Technical Report, AFRL-IF-RS, http://stinet.dtic.mil.

Yue, S.T., Singer, V.L., Roth, B.L., Mozer, T.J., Millard, P.J., Jones, L.J., Jin, X. and Haugland, R.P. (1997). Substituted unsymmetrical cyanine dyes with selected permeability, US,patent 5658751.

Zipper, H., Lämmle, K., Buta, C., Brunner, H., Bernhagen, J., Vitzthum, F. Investigations on the binding of SYBR Green I to double-stranded DNA, Proceedings, German Society for Biochemistry and Molecular Biology & German Society for Experimental and clinical Pharmacology and Toxicology, Halle, Germany, Sept. 7-10-2002.

Zipper, H., Brunner, H., Bernhage, J., Vitzthum, F. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. Nucleic Acids Res. 2004, 32(12), e103.

Zucker, M., Matthews, D.H., Turner, D.H. (1999) Algorithms and thermodynamics for RNA secondary sructure prediction: a practical guide. Braciszewski, J. and Clark, B.F.C. (eds) In *RNA Biochemistry and Biotechnology*, NATO ASI Series, Kluwer Academic Publishers.

Zucker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 9, 133-148.